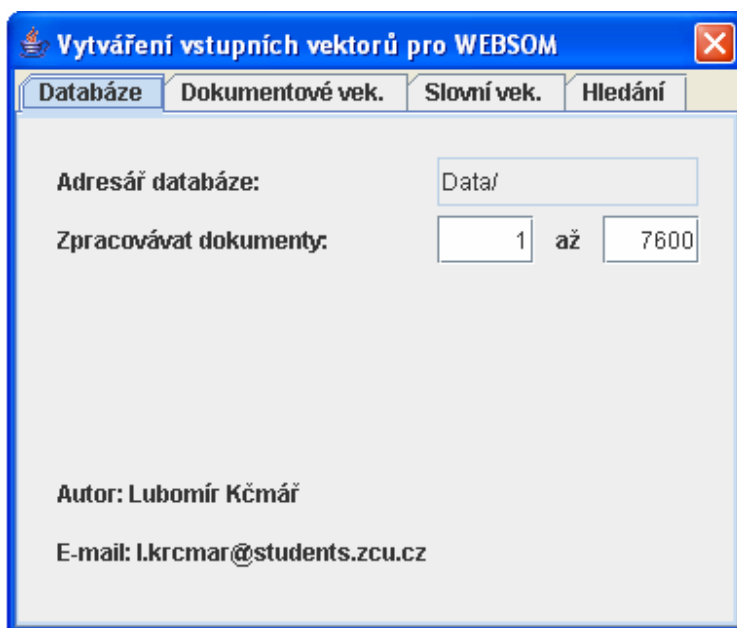


## Ovládání aplikace

Ovládání aplikace je velmi jednoduché. Aplikace sestává ze čtyř záložek. První záložkou je databáze. Následují záložky Dokumentové vektory, Slovní vektory a Hledání.

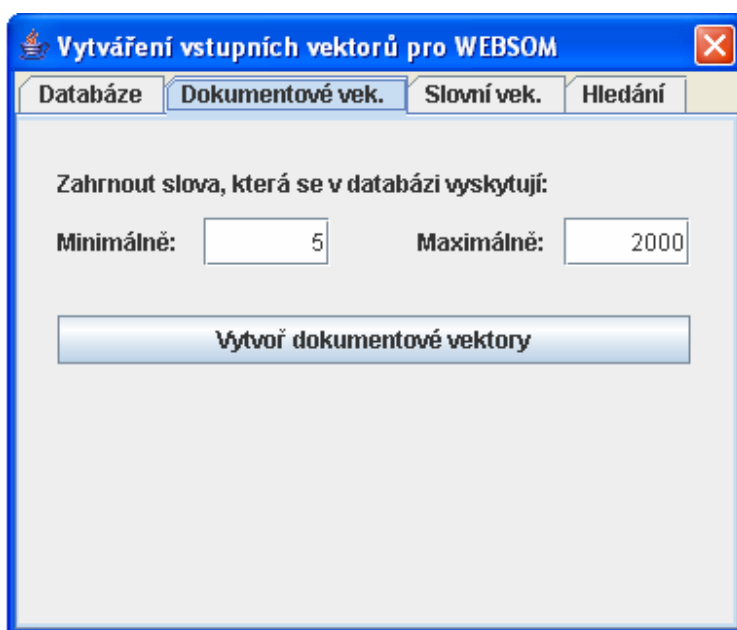
První záložka umožňuje volbu počtu zpracovávaných dokumentů. Dokumenty ČTK jsou číslovány od 1 do 7600. První hodnota udává první zpracovávaný dokument. Druhá hodnota udává dokument, který se bude zpracovávat jako poslední.



The screenshot shows a Windows-style application window titled "Vytváření vstupních vektorů pro WEBSOM". It has four tabs: "Databáze", "Dokumentové vek.", "Slovní vek.", and "Hledání". The "Databáze" tab is selected. Inside the window, there are two labels with corresponding input fields: "Adresář databáze:" with a text box containing "Data/" and "Zpracovávat dokumenty:" with two numeric input boxes containing "1" and "7600" separated by the word "až". At the bottom, there is contact information: "Autor: Lubomír Křmář" and "E-mail: l.krcmar@students.zcu.cz".

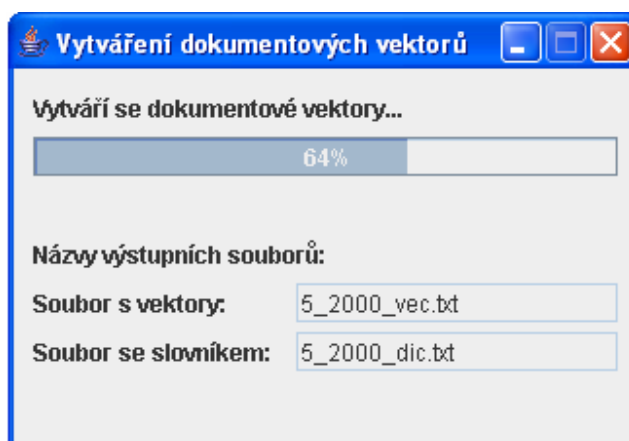
Obr. 33: První záložka

Druhá záložka slouží pro vytváření dokumentových vektorů k části databáze vymezené v první záložce. Aplikace umožňuje zadávání minimální a maximální četnosti slov, ze kterých bude vytvořen slovník. Slova, která se budou v celé databázi vyskytovat méněkrát či vícekrát, nebudou ve slovníku. Nebudou tak tvořit položky ve vytvořených dokumentových vektorech. Proces vytváření dokumentových vektorů pro zadané parametry se spouští stiskem tlačítka Vytvoř dokumentové vektory.



Obr. 34: Druhá záložka

Po stisknutí tlačítka dojde k zobrazení informačního okénka o průběhu vytváření dokumentových vektorů.



Obr. 35: Informační okénko pro vytváření dokumentových vektorů

Třetí záložka slouží pro vytváření slovních vektorů k části databáze vymezené v první záložce. Aplikace umožňuje zadávání minimální četnosti slov, ze kterých bude vytvořen slovník. Slova, která se budou v celé databázi vyskytovat méněkrát, nebudou ve slovníku. Nebude k nim tak vytvořen žádný slovní vektor. Nastavitelným parametrem je také délka slovního vektoru (výsledný slovní vektor bude trojnásobný). Dalším nastavitelným parametrem je parametr epsilon (konstanta vytvořená podle finského vzoru). Ve třetí záložce je možné nastavit, budou-li se vytvářet dokumenty s nahrazenými slovy za slovní vektory a budou-li se vytvářet seznamy slov následujících a předcházejících ke slovům, ke kterým se vytváří slovní vektor. Proces vytváření slovních vektorů pro zadané parametry se spouští stiskem tlačítka Vytvoř slovní vektory.

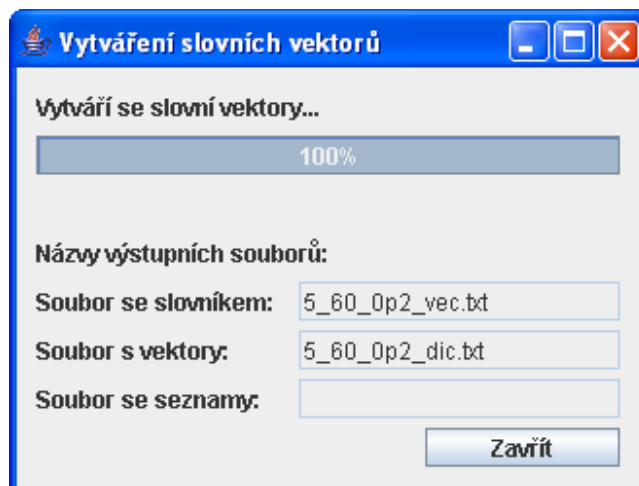
The screenshot shows a Windows-style dialog box titled "Vytváření vstupních vektorů pro WEBSOM". It has four tabs: "Databáze", "Dokumentové vek.", "Slovní vek." (which is selected), and "Hledání". The "Slovní vek." tab contains the following settings:

- Minimální počet výskytů slova v databázi:** 5
- Délka vektoru ke slovu:** 20
- Délka vektoru ke slovu po zpracování:** 60
- Parametr epsilon:** 0.2
- ☐ Nahradiť slova dokumentů vektory
- ☐ Vytvořit seznamy slov předcházejících a následujících

At the bottom of the dialog is a large button labeled "Vytvoř slovní vektory".

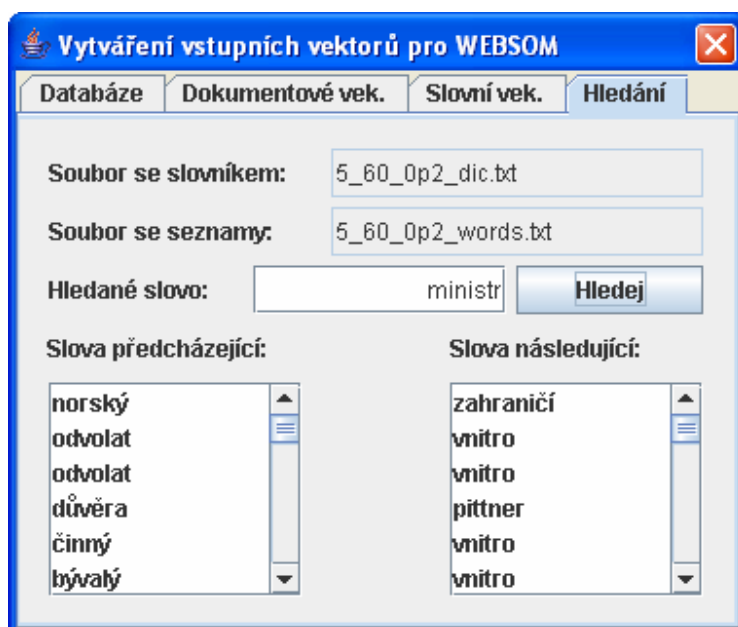
Obr. 36: Třetí záložka

Po stisknutí tlačítka dojde k zobrazení informačního okénka o průběhu vytváření slovních vektorů. V případě zaškrtnutého políčka Předělat slova dokumentů na vektory bude ve čtvrté záložce zpřístupněna možnost vyhledávání.



Obr. 37: Informační okénko pro vytváření slovních vektorů

Čtvrtá záložka slouží pro vyhledávání slov předcházejících a následujících ke slovům, pro které byly v předchozím spuštěném procesu vytvořeny slovní vektory. Seznamy jsou slova, ze kterých byly slovní vektory vytvořeny. Zadávaným parametrem je pouze hledané slovo. Proces vyhledávání seznamů se spouští stiskem tlačítka Hledej.



Obr. 38: Čtvrtá záložka

Výstupem procesů vytváření dokumentových a slovních vektorů jsou soubory. V názvech souborů jsou uloženy parametry, s kterými byly vytvářeny. Soubory jsou hierarchicky ukládány podle nastavených parametrů.

Příkladem hierarchického uložení výstupních souborů pro proces Vytváření dokumentových vektorů, s parametry databáze 1, 1200 a s parametry v záložce Dokumentové vektory 5, 2000, jsou soubory s názvy 5\_2000\_dic.txt (slovník) a 5\_2000\_vec.txt (vektory). Soubory jsou uloženy v adresáři s názvem DoVe\_1\_1200, který se nachází ve stejném adresáři, ze kterého byla spuštěna aplikace.